UNIVERSITÄT BERN

Faculty of Business, Economics and Social Sciences

Department of Economics

Detailed Decompositions in Generalized Linear Models

Boris Kaiser

13-09

October 2013

DISCUSSION PAPERS

Schanzeneckstrasse 1
Postfach 8573
CH-3001 Bern, Switzerland
http://www.vwi.unibe.ch

Detailed Decompositions in Generalized Linear Models

Boris Kaiser*

October 28, 2013

Abstract
We propose a new approach for performing detailed decompositions of average outcome differentials, which can be applied to all types of generalized linear models. A simulation exercise demonstrates that our method produces more convincing results than existing methods. An empirical application to the immigrant-native wage differential in Switzerland is presented.

Keywords: Oaxaca-Blinder; Detailed Decomposition; Generalized Linear Models

JEL: C10; C50; C51; J31

*Department of Economics, Schanzeneckstrasse 1, University of Bern, CH-3001 Bern, Switzerland. E-mail: boris.kaiser@vwi.unibe.ch. Phone: +41 31 631 40 49.

1 Introduction

Decomposition methods are frequently employed to analyze differences in average economic outcomes between groups of individuals. The most prominent example is the decomposition in the linear model by Oaxaca (1973) and Blinder (1973) which splits the gap into a structural effect (due to differences in coefficients) and a composition effect (due to differences in covariates). Since it is usually of interest how strongly individual explanatory variables drive the differential, these aggregate effects must be further broken down by a so-called detailed decomposition.

While detailed decompositions are straightforward in the linear Oaxaca-Blinder framework, they are more complicated in nonlinear models. A major drawback of some approaches suggested in the literature is that the detailed decomposition is path-dependent, i.e. sensitive to the order of computation (e.g. Fairlie, 2005). Another approach, which does not suffer from path-dependence, is proposed by Yun (2004, 2008), who approximates the nonlinear decomposition terms at the sample means of the covariates with a first-order Taylor approximation. However, the linear approximation has the disadvantage that the decomposition neglects the effects of covariate distributions on the nonlinear outcome models.

In this paper, we propose a new approach for performing detailed decompositions which takes into account the effects of all higher-order moments of covariates on the outcome model. The method is path-independent and the contributions of individual covariates add up to the aggregate decomposition, both of which are desirable properties. Moreover, our detailed decomposition generalizes to the Oaxaca-Blinder case in the linear model. The method we suggest can be applied to any generalized linear model (GLM), a framework which encompasses numerous limited dependent variable models as well as the linear model.

A small simulation study demonstrates that our procedure produces more appropriate results than Yun's method if covariates differ in higher-order moments. Furthermore, we provide an empirical application to the native-immigrant mean wage gap in Switzerland.

The remainder of this paper is organized as follows: Section 2 develops the framework for the detailed decomposition proposed in this paper. Section 3 provides the simulation exercise and Section 4 the empirical application. Finally, Section 5 concludes.

2 Decomposition Framework

The framework is based on potential-outcomes notation (Rubin, 1974) as is standard in the modern approach to decomposition analysis (Firpo et al., 2011). We consider a population of two non-overlapping groups denoted groups A and B, which for example, could represent male and female workers. If individual $i \in \mathcal{A}$, we have $D_i = A$ and if $i \in \mathcal{B}$, we have $D_i = B$. Denote y_{iB} the potential outcome of individual i if she belongs to group B and y_{iA} the potential outcome if she belongs to group A. The observed outcome of person i is thus $y_i = \mathbb{I}(D_i = B)y_{iB} + \mathbb{I}(D_i = A)y_{iA}$. Hence, y_{iB} for units in group A and y_{iA} for units in group B are counterfactual.

As for any decomposition problem, a counterfactual outcome of interest must be defined.[1] As in Firpo et al. (2011), we focus on the "simple" counterfactual $E[y_{iA}|D_i = B]$,

[1] In the classic example of Oaxaca-Blinder decomposition of wages, the counterfactual outcome of interest is implicitly set by the choice of the reference wage structure (Cotton, 1988; Neumark, 1988; Jann, 2008).

which corresponds to the mean outcome in group B that would prevail if their outcome distribution were generated by the data generating process of group A.[2] Given this counterfactual, the aggregate decomposition of average outcomes can be written as

$$\Delta = (\underbrace{E[y_i|D_i = B] - E[y_{iA}|D_i = B]}_{=\Delta^S}) + (\underbrace{E[y_{iA}|D_i = B] - E[y_i|D_i = A]}_{=\Delta^X}), \tag{1}$$

where $E[y_{iA}|D_i = A]$ and $E[y_{iB}|D_i = B]$ have been replaced by the corresponding observed outcomes. The term Δ^S is the **structural effect** (also: coefficients effect), which corresponds to the average treatment effect on the treated (Imbens and Wooldridge, 2009). The term Δ^X is the **composition effect** (also: characteristics effect) and captures the part of the gap due to differences in covariates across groups.

The aggregate decomposition in (1) is nonparametrically identified under the well-known ignorability assumptions, but additional assumptions are necessary to identify the contributions of individual covariates in a detailed decomposition (Firpo et al., 2011).[3] To allow for practical and feasible estimation, we consider a fully parametric framework. That is, we assume that the conditional expectation functions (CEF) take the form of a generalized linear model (GLM)

$$E[y_i|X_i, D_i = g] = G(X_i\beta^g) \quad \text{for } g = \{A, B\}, \tag{2}$$

where $\beta^g \in \mathbb{R}^k$ is a unique column vector of population coefficients and X_i is a $(1 \times K)$ row vector of covariates with the first element being a constant. We assume that X_i contains at least one continuous covariate for reasons explained below. The central element of any GLM is the function $G(\cdot)$, which is a monotonic and differentiable link function that maps the linear index $X_i\beta^g$ one-to-one into the support of the outcome. By the law of iterated expectations, the aggregate decomposition in (1) can now be written as

$$\begin{aligned} \Delta = &\underbrace{E[G(X_i\beta^B)|D_i = B] - E[G(X_i\beta^A)|D_i = B]}_{\Delta^S} \\ &+ \underbrace{E[G(X_i\beta^A)|D_i = B] - E[G(X_i\beta^A)|D_i = A]}_{\Delta^X} \end{aligned} \tag{3}$$

The aim of a detailed decomposition is to further break down the terms Δ^X and Δ^S into the contributions of individual covariates.

2.1 Detailed Decomposition

In both linear and nonlinear models, detailed decompositions of the structural effect do not have a meaningful interpretation for covariates without a natural zero point (Oaxaca and Ransom, 1999; Jann, 2008; Firpo et al., 2011).[4] Due to these practical limitations, we will confine the discussion to the decomposition of the composition effect, which does

[2]Note that our decomposition approach does not depend on the choice of counterfactual. We merely choose it because it is easy to interpret and has a meaningful treatment-effects equivalent.

[3] Ignorability states that (i) potential outcomes are independent of group assignment given covariates and (ii) the covariate support of group B is contained within the covariate support of group A. Ignorability implies $E[y_{iA}|D_i = B] = E_X\{E[y_i|X_i, D_i = A]|D_i = B\}$.

[4]Yun (2005) proposes a solution to the problem consisting of an ex-post normalization on the coefficients. However, as Firpo et al. (2011) note, such normalizations come at the cost of interpretability.

not suffer from this problem. Appendix A briefly sketches how the structural effect can be decomposed with our method.

While detailed decompositions in nonlinear models are less straightforward than in linear models (because contributions of covariates are not additively separable), several approaches have been suggested in the literature. First, Fairlie (2005) applies a simulation procedure to binary probability models, but the resulting decomposition is sensitive to the order of computation (path-dependent). Second, Yun (2004) proposes a path-independent procedure which is derived from two approximation steps. First, he approximates the aggregate decomposition terms at the sample means, and second, he performs a first-order Taylor approximation. The resulting expressions are covariate "weights" in which all terms containing $G(\cdot)$ cancel out. The drawback of this detailed decomposition is that it neglects the impact of higher-order moments of X_i on the nonlinear model and only takes into account differences in means. This is relevant because, unlike in linear models, average outcomes in nonlinear model are affected, for example, by different covariate variances across groups. In what follows, we present an alternative approach that does not have this limitation.

The main difference to Yun's approach is that we start by defining the detailed decomposition terms *conditional on covariates.* In this way, there is no need to approximate the aggregate decomposition terms at the sample means in the first place. To decompose the composition effect in a meaningful way, the contributions of individual covariates can be related to a counterfactual experiment. Given $i \in \mathcal{A}$, an arbitrary observation from group A, we want to measure the importance of covariate k in the change of the CEF value when covariates are switched from the observed values X_i to some counterfactual values, denoted by X_i^c. After having defined the *conditional* decomposition terms from this experiment, we will then be able to obtain *unconditional* decomposition terms in a second step.

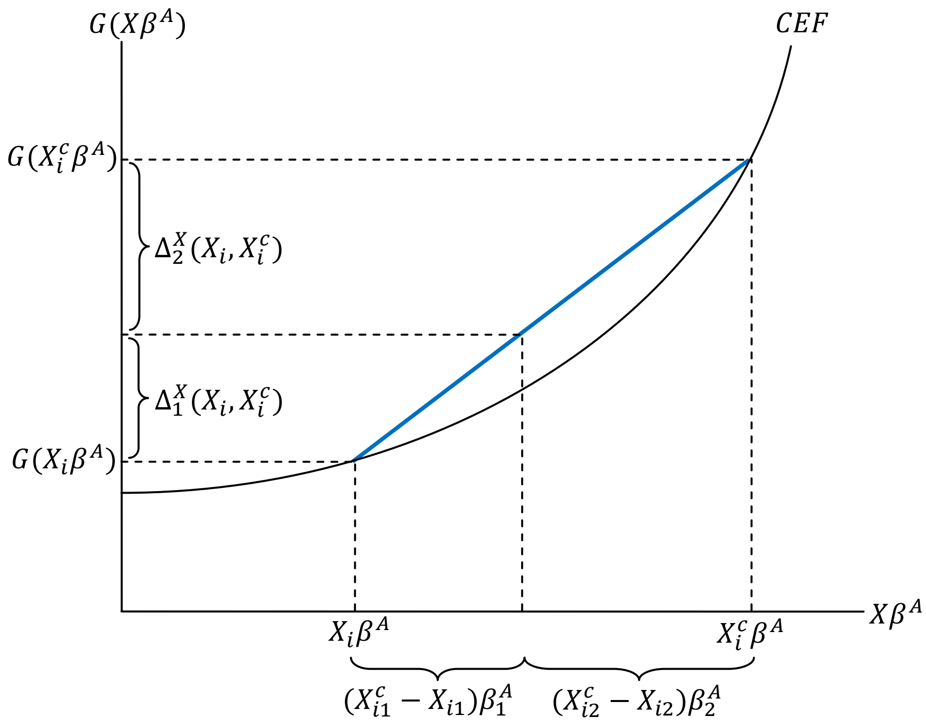
The main idea of our conditional detailed decomposition is to decompose the distance between the two points on the CEF, i.e. between $G(X_i^c\beta^A)$ and $G(X_i\beta^A)$. But in order to ensure path independence, the conditional decomposition we suggest is based on the line segment between the two points in the two dimensional space of the CEF and the linear index. The contribution of an individual covariate k is then defined by the slope of this line segment times $(X_{ik}^c - X_{ik})\beta_k^A$. More formally, we define the contribution of covariate k when we move from the observed outcome to the counterfactual outcome as

$$\Delta_k^X(X_i, X_i^c) = \frac{G(X_i^c\beta^A) - G(X_i\beta^A)}{(X_i^c - X_i)\beta^A}(X_{ik}^c - X_{ik})\beta_k^A \tag{4}$$

where the first term is the slope of the line segment and the second term is the horizontal distance in the linear index attributable to covariate k. Figure 1 provides a graphical illustration of the approach for a simple example with two covariates and a convex CEF. It is clear from Figure 1 that the order in which we compute the decomposition terms does not matter (path-independence) and that the two detailed decomposition terms, $\Delta_1^X(X_i, X_i^c)$ and $\Delta_2^X(X_i, X_i^c)$, add up to the total difference $\Delta^X(X_i, X_i^c) = G(X_i^c\beta^A) - G(X_i\beta^A)$.

Note that $P[(X_i^c - X_i)\beta^A = 0] = 0$ due to the presence of continuous covariates such that (4) is always well-defined. In the case where $(X_i^c - X_i)\beta^A$ gets arbitrarily close to zero, it follows immediately that $\frac{G(X_i^c\beta^A) - G(X_i\beta^A)}{(X_i^c - X_i)\beta^A} \to G'(X_i\beta^A)$. Thus, (4) remains well-defined because $G(\cdot)$ is a differentiable function. In the special case where all covariates

Figure 1: Conditional Detailed Decomposition



are discrete, (4) can be slightly modified.[5]

Now we turn to the choice of X_i^c. Given our decomposition framework, it is natural that the counterfactual covariate distribution of interest is the covariate distribution of group B. That is, we are interested in an experiment in which we set $X_i^c = X_j$ where $j \in \mathcal{B}$. The term $\Delta_k^X(X_i, X_j)$ then measures the importance of covariate k in the change of the CEF when observation i from group A is assigned the characteristics of observation j from group B.

Now the issue remains to go from a conditional detailed decomposition $(\Delta_k^X(X_i, X_j))$ to an unconditional detailed decomposition (Δ_k^X). We obtain the latter in two steps. In the first step, we obtain $\Delta_k^X(X_i, X_j)$ for a given unit $i \in \mathcal{A}$ and the entire population $\mathcal{B}$. These terms are appropriately averaged, such that we obtain the average contribution of covariate k *conditional on* X_i, i.e. $\Delta_k^X(X_i)$. This exercise is performed for the entire population $\mathcal{A}$. In the second step, the conditional contributions obtained in step one are then simply averaged across population $\mathcal{A}$, which yields the unconditional contribution of covariate k to the composition effect, i.e. Δ_k^X. Putting it more formally, the unconditional contribution can be obtained by integrating $\Delta_k^X(X_i, X_j)$ over the covariate distributions

[5]If all covariates are discrete, it may be that $P[(X_i^c - X_i)\beta^A = 0] \neq 0$. In this case we define the contribution of covariate k as

$$\Delta_k^X(X_i, X_i^c) = \begin{cases} \frac{G(X_i^c\beta^A) - G(X_i\beta^A)}{(X_i^c - X_i)\beta^A}(X_{ik}^c - X_{ik})\beta_k^A & \text{if } (X_i - X_i^c)\beta^A \neq 0 \\ G'(X_i\beta^A)(X_{ik}^c - X_{ik})\beta_k^A & \text{if } (X_i - X_i^c)\beta^A = 0 \end{cases} \quad (5)$$

in the two sub-populations:

$$\Delta_k^X = \int_u \int_v \Delta_k^X(u,v) dF_{X|D=B}(v) dF_{X|D=A}(u) \tag{6}$$

Due to the linearity of the operator, it follows immediately that the unconditional contributions of all covariates sum up to the aggregate composition effect, $\Delta^X = \sum_{k=1}^K \Delta_k^X$. Given a finite sample with N_A units in group A and N_B units in group B, the corresponding estimator is given by the sample analogue:

$$\widehat{\Delta}_k^X = \frac{1}{N_A} \sum_{i:D_i=A} \frac{1}{N_B} \sum_{j:D_j=B} \Delta_k^X(X_i, X_j). \tag{7}$$

This estimator belongs to the class of two-sample U-statistics (as for example the Mann-Whitney-Wilcoxian rank sum test) and is shown to be consistent and asymptotically normal under standard regularity conditions and given that $N_A/(N_A+N_B)$ is bounded away from zero and infinity (see e.g. Lee and Dehling, 2005, for a formal proof).

The proposed procedure for the detailed decomposition has some attractive properties. First, as shown, the individual contributions add up to the aggregate decomposition terms. Second, the decomposition is path-independent, i.e. does not depend on the order of computation (as opposed to Fairlie's (2005) method). Third, it is easy to see that if CEFs are linear, the decomposition terms reduce to the standard detailed decomposition in the linear model. In other words, if $G(\cdot)$ is the identity function $(\cdot)$, (7) reduces to $\widehat{\Delta}_k^X = (\bar{X}_k^B - \bar{X}_k^A)\hat{\beta}_k^A$, where $\bar{X}_k^B$ and $\bar{X}_k^A$ are group-specific sample means. Thus, our method can be regarded as a natural extension of the standard detailed decomposition to nonlinear settings.

Finally, we compare our approach to Yun's (2004) from a statistical viewpoint. He performs two approximations directly to the *unconditional* composition effect, which yields the following detailed decomposition terms:

$$\Delta_{k,approx}^X = \Delta^X \cdot \frac{(E[X_{jk}|D_j = B] - E[X_{ik}|D_i = A])\beta_k^A}{(E[X_j|D_j = B] - E[X_i|D_i = A])\beta^A}. \tag{8}$$

Yun's (2004) approach therefore amounts to integrating all terms containing X_i *separately* as a result of the approximation. In contrast, our method integrates all terms in (4) *jointly* across the relevant populations. In this way, our method accounts for group-specific differences in the higher-order moments of X_i and not only for differences in means as in (8).

2.2 Estimation

In practice, of course, the GLM must be estimated in a first step to obtain estimates of the parameters (β^A, β^B). A GLM assumes that the dependent variable is generated from a probability distribution in the exponential family (normal, Poisson, gamma, Bernoulli, binomial, categorical, multinomial, etc.). However, a well-known result from Gourieroux et al. (1984) states that consistent estimation of the parameters only requires correct specification of the CEF and no further distributional assumptions. These estimators are called quasi-maximum-likelihood estimators because they are consistent even if the underlying density is misspecified.[6]

[6] The most famous example for a QML estimator is OLS: it corresponds to the ML estimator derived under normality and is consistent even if errors are not normal.

Due to the two-step nature of the estimation of the detailed decomposition, the bootstrap can be used to compute standard errors and conduct inference. In this context, one drawback of our method is that estimation of Δ_k^X can become computationally demanding if the sample size is large.[7]

3 Simulation Exercise

We conduct a small simulation study to compare the detailed decomposition proposed in this paper with the method of Yun (2004). For the purpose of illustration, we consider a very simple setup with only two covariates and zero structural effect (i.e. $\Delta = \Delta^X$). Both covariates are normally distributed. The outcome is nonlinear in the covariates and generated by an exponential CEF, i.e. $y = \exp(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2)$, where we set $\alpha_0 = 0$ and $\alpha_1 = \alpha_2 = 1$. We keep the model deterministic because we are only interested in the detailed decomposition and not in the estimation of the parameters.[8] We generate random samples of size 2000, with $N_A = N_B = 1000$.

In the first experiment, we generate a gap by differences in means only. In group B, the means of both covariates are larger than the means in group A, but variances are equal, i.e. $x_k|D = A \sim N(1, 1)$ and $x_k|D = B \sim N(2, 1)$ for $k \in \{1, 2\}$. In this example, the "correct" detailed decomposition is unambiguous; we expect that the contributions of the two covariates to the composition effect are equal. Table 1 shows the results from this first experiment and clearly confirms our expectations. As we can see, both

Table 1: Monte-Carlo Simulation Results, Experiment 1

	configuration of covariates			simulation results			
				our method		Yun's method	
	group A	group B		mean	st.dev.	mean	st.dev.
x_1	$N(1,1)$	$N(2,1)$	Δ_1^X	64.23	6.51	64.21	6.42
x_2	$N(1,1)$	$N(2,1)$	Δ_2^X	64.26	6.45	64.28	6.42
			total gap (Δ^X)	128.49	12.22	128.49	12.22

Notes: 1000 replications of samples with $N_A = 1000$ and $N_B = 1000$.

decompositions attribute the same amount of the total gap to either covariate because there are only differences in first moments and these differences are the same for both covariates. Note also that by construction, individual contributions of covariates add up to the total composition effect in both methods.

In the second experiment, x_2 is the same as before but we add a mean-preserving spread to x_1 in that we increase the variance of x_1 in group B by a factor of 4, i.e. $x_1|D = B \sim N(2, 4)$, while keeping everything else unchanged. Due to the convexity of the outcome model, the large dispersion of x_1 in group B sizably increases the average outcome of this group. As a result of the increased variance, the total gap increases by a factor of about 5. Thus, covariate x_1 is now much more important in explaining the

[7]Appendix B demonstrates how the detailed decomposition proposed above can be computed efficiently using matrix algebra. The decomposition method is implemented in the STATA package `glmdeco` which is available on the author's homepage at http://staff.vwi.unibe.ch/kaiser/research.html

[8]Note that adding an error term does not affect the simulation results in a meaningful way because it only adds noise to the outcome model.

composition effect which should be reflected in a larger contribution of x_1 relative to the contribution of x_2. Table 2 shows the simulation results from the second experiment. On

Table 2: Monte-Carlo Simulation Results, Experiment 2

	configuration of covariates			simulation results our method		Yun's method	
	group A	group B		mean	st.dev.	mean	st.dev.
x_1	$N(1,1)$	$N(2,4)$	Δ_1^X	461.23	186.93	322.57	128.09
x_2	$N(1,1)$	$N(2,1)$	Δ_2^X	183.46	64.74	322.12	121.41
			total gap (Δ^X)	644.69	247.91	644.69	247.91

Notes: 1000 replications of samples with $N_A = 1000$ and $N_B = 1000$.

average, Yun's method still attributes about half of the gap to x_1 and x_2, respectively, because the effects of differences in variances are not taken into account. In contrast, our decomposition method produces results that are in line with the reasoning above: Δ_1^X is considerably larger than Δ_2^X, i.e. the biggest share of the gap is attributed to x_1. This is because the differences in the variance are taken into account by our decomposition method. Note that Δ_2^X is also larger in absolute value relative to the first experiment in Table 1. The reason is that the *joint distribution* of (x_1, x_2), which is different in the two experiments, affects the detailed decomposition, as is evident from equation (6).

We conduct a third simulation experiment, in which we want to assess the precision of the two estimators when differences in group-specific covariate distributions become small. For this purpose, we modify the first experiment by reducing differences in covariate means to 0.05, while leaving everything else the same. Simulation results are presented in Table 3. Most notably, we find that the estimator of Yun's decomposition behaves

Table 3: Monte-Carlo Simulation Results, Experiment 3

	configuration of covariates			simulation results our method		Yun's method	
	group A	group B		mean	st.dev.	mean	st.dev.
x_1	$N(1,1)$	$N(1.05,1)$	Δ_1^X	1.022	1.366	1.550	10.745
x_2	$N(1,1)$	$N(1.05,1)$	Δ_2^X	1.067	1.372	0.539	10.860
			total gap (Δ^X)	2.089	2.365	2.089	2.365

Notes: 1000 replications of samples with $N_A = 1000$ and $N_B = 1000$.

extremely erratically as reflected in the very large standard deviations. A density plot of the distribution of Monte Carlo estimates (not shown) reveals that the estimator follows a non-normal, degenerate distribution with fat tails. In contrast, the estimator of our decomposition method remains well-behaved; the distribution of Monte Carlo estimates (not shown) is approximately normal as implied by the theoretical results of Lee and Dehling (2005).

4 Application

To illustrate the detailed decomposition method, we provide an empirical application on the *arithmetic* mean wage gap between native and immigrant workers. While researchers commonly decompose the approximate *geometric* mean wage gap (i.e. the log-wage gap $E[\ln y_i|D_i = B] - E[\ln y_i|D_i = A]$) using a log-linear regression model, the *arithmetic* mean wage gap, $E[y_i|D_i = B] - E[y_i|D_i = A]$, might be the more appropriate quantity (Leslie and Murphy, 1997; Blackburn, 2008). This argument suggests that one should model the original dependent variable (raw wage) directly. The standard functional form assumption then implies that the wage should be specified as an exponential function of the covariates. This model can be estimated with quasi-maximum-likelihood techniques (Gourieroux et al., 1984).

The data is drawn from the Swiss Earnings Structure Survey in 2008. This data is of higher quality than comparable survey data because it is elicited directly from employers' records. Moreover, the large sample size ensures that sampling bias is not a concern. For the purpose of illustration, we confine the analysis to male full-time workers in German-speaking Switzerland.[9] Omitting observations with missing values and those aged under 20 or above 65, the sample contains more than 400,000 observations. Immigrant workers are defined as those without Swiss citizenship. The outcome of interest is full-time equivalent gross monthly earnings[10] and the set of controls consists of educational attainment (9 categories), potential work experience in years (quadratic), tenure in years (quadratic) and marital status (3 categories).

Table 4 presents descriptive statistics. Immigrants constitute about 30% of adult male employment in German-speaking Switzerland and earn considerably lower wages on average. Comparing covariates, we see, for example, that natives have more work experience and tenure and the dispersion of these covariates is also larger. Given the convexity of the assumed CEF, we would therefore expect that Yun's decomposition attributes too small a share of the composition effect to these variables.

For the decomposition analysis, we define natives as group A and immigrants as group B, such that the the counterfactual of interest in the mean wage immigrants would be paid if their wages had been generated by the wage structure of natives. Table 5 presents the results from the detailed decomposition across the two decomposition methods.[11] As we can see, about half of the observed differential is explained by different covariate distributions. The detailed decomposition explores the extent to which individual covariates contribute to the mean wage gap. We observe some differences between the two methods compared. For example, while the method of Yun (2004) suggests that tenure does not significantly affect the gap, our method suggests that it widens the gap significantly on a 0.1% level. Conversely, the importance of educational attainment is smaller when our method is used. To summarize, we find that it matters in practice for a detailed decomposition in a nonlinear model whether we allow the entire covariate distributions (or only their means) to affect the detailed decomposition terms. In our application, we find that both the quantitative and qualitative results are different.

[9] "Full-time" are those whose hours worked are at least 90% of a full-time equivalent.

[10] Including monetary benefits, extra pay for night-shifts or weekend-shifts and, where applicable, one twelfth of the 13th monthly salary.

[11] We report the decomposition of the structural effect for the sake of completeness, but results are only presented for continuous covariates for the reasons mentioned previously.

Table 4: Descriptive Statistics

	natives mean	immigrants mean	diff.
monthly gross wage (in CHF)	7277.015	6421.693	-855.322
	(4845.426)	(4949.802)	
log(wage)	8.788	8.652	-0.136
	(0.411)	(0.420)	
education level			
university	0.058	0.083	0.026
college	0.069	0.046	-0.024
higher vocational training	0.149	0.065	-0.084
teaching diploma	0.003	0.002	-0.001
secondary school	0.016	0.013	-0.002
vocational training	0.619	0.419	-0.200
firm-specific vocational training	0.020	0.071	0.051
primary school	0.047	0.209	0.162
other education	0.021	0.093	0.072
marital status			
single	0.380	0.309	-0.071
married	0.540	0.625	0.084
divorced, widowed	0.080	0.067	-0.013
work experience (years)	22.467	20.946	-1.521
	(11.733)	(10.513)	
tenure (years)	9.772	6.982	-2.790
	(9.852)	(7.902)	
# observations	315192	134293	

Notes: The sample consists of male full-time workers from German-speaking Switzerland in 2008. Standard deviations are in parentheses. Sampling weights are used. *Source*: Swiss Wage Structure Survey, Swiss Federal Statistical Office.

5 Conclusions

This paper has presented a new approach for performing detailed decompositions to differences in means when the outcome model is nonlinear and belongs to the class of generalized linear models (GLM). As opposed to Yun's (2004) method, we derive a conditional detailed decomposition which takes into account that differences in higher-order moments of covariates affect average outcomes through the nonlinearity of the model. A simulation exercise has demonstrated that our method produces more convincing results when the dispersion of covariates differs across groups. Furthermore, the analysis of the immigrant-native wage gap exemplifies that these effects can be relevant in empirical applications.

A potential area of future research could be to exploit the proposed procedure for decompositions of more general distributional statistics such as variances, quantiles, etc. To do this, our method could be combined with the distribution regression approach recently

Table 5: Detailed Decomposition of Immigrant-Native Wage Gap

	our method			Yun's method		
	est.	st.err.	in %	est.	st.err.	in %
total differential	-855.3^{***}	(21.0)	100.0	-855.3^{***}	(19.5)	100.0
aggregate composition effect	-414.7^{***}	(14.3)	48.5	-414.7^{***}	(14.6)	48.5
education level	-329.6^{***}	(12.6)	38.5	-376.3^{***}	(12.6)	44.0
work experience (years)	45.8^{***}	(2.1)	-5.4	53.4^{***}	(2.2)	-6.2
tenure (years)	-27.7^{***}	(4.9)	3.2	-1.3	(4.8)	0.1
marital status	-103.2^{***}	(3.8)	12.1	-90.6^{***}	(3.4)	10.6
aggregate structural effect	-440.6^{***}	(19.1)	51.5	-440.6^{***}	(19.3)	51.5
work experience (years)	-28.8	(17.9)	3.4	-17.2	(16.9)	2.0
tenure (years)	-155.5^{*}	(69.5)	18.2	-135.4^{*}	(55.2)	15.8
# observations $D_i = 1$ (immigrants): 134293						
# observations $D_i = 0$ (natives): 315192						

Notes: Estimation of outcome model is based on Poisson Quasi-Maximum-Likelihood. The sample consists of male full-time workers from German-speaking Switzerland in 2008. Standard errors are bootstrapped using 1000 draws of 10%-subsamples. Sampling weights are used. Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

studied by Chernozhukov et al. (2013, forthcoming), which is based on the estimation of binary probability models such as logit or probit.

6 Bibliography

Blackburn, M.K.L., "Are Union Wage Differentials in the United States Falling?," *Industrial Relations: A Journal of Economy and Society*, 2008, *47* (3), 390–418.

Blinder, Alan S., "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 1973, *8* (4), 436–455.

Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly, "Inference on counterfactual distributions," *Econometrica*, 2013, forthcoming.

Cotton, Jeremiah, "On the decomposition of wage differentials," *The Review of Economics and Statistics*, 1988, *70* (2), 236–243.

Fairlie, Robert W, "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models," *Journal of Economic and Social Measurement*, 2005, *30* (4), 305–316.

Firpo, Sergio, Nicole Fortin, and Thomas Lemieux, "Decomposition methods in economics," *Handbook of Labor Economics*, 2011, *4*, 1–102.

Gourieroux, Christian, Alain Monfort, and Alain Trognon, "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 1984, *52* (3), 681–700.

Imbens, Guido W. and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 2009, *47* (1), 5–86.

Jann, Ben, "The Blinder-Oaxaca decomposition for linear regression models," *Stata Journal*, December 2008, *8* (4), 453–479.

Lee, Mei-Ling Ting and Herold G. Dehling, "Generalized two-sample U-statistics for clustered data.," *Statistica Neerlandica*, 2005, *59* (3), 313–323.

Leslie, Derek and Phil Murphy, "Measuring discrimination by decomposing earnings functions," *Applied Economics Letters*, 1997, *4* (2), 117–120.

Neumark, David, "Employers' discriminatory behavior and the estimation of wage discrimination," *Journal of Human Resources*, 1988, *23* (3), 279–295.

Oaxaca, Ronald L., "Male-female wage differentials in urban labor markets," *International Economic Review*, 1973, *14* (3), 693–709.

_ **and Michael R. Ransom**, "Identification in detailed wage decompositions," *Review of Economics and Statistics*, 1999, *81* (1), 154–157.

Rubin, Donald B, "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of Educational Psychology*, 1974, *66* (5), 688–701.

Yun, Myeong-Su, "Decomposing differences in the first moment," *Economics letters*, 2004, *82* (2), 275–280.

_ , "A simple solution to the identification problem in detailed wage decompositions," *Economic Inquiry*, 2005, *43* (4), 766–772.

_ , "Identification problem and detailed Oaxaca decomposition: a general solution and inference," *Journal of Economic and Social Measurement*, 2008, *33* (1), 27–38.

Appendix

A Decomposition of the Structural Effect

If researchers are willing to decompose the structural effect despite the interpretational difficulties (Oaxaca and Ransom, 1999), the contribution of covariate k to the structural effect conditional on X_j, where $j \in \mathcal{B}$, is given by

$$\Delta_k^S(X_j) = \frac{G(X_j\beta^B) - G(X_j\beta^A)}{X_j(\beta^B - \beta^A)} X_{jk}(\beta_k^B - \beta_k^A)$$

In the case where all covariates are discrete, a similar modification as the one for the composition effect is made (see Section 2). Integrating this expression over the covariate distribution of group B, we obtain the unconditional contribution of covariate k to the structural effect

$$\Delta_k^S = \int_u \Delta_k^S(u) dF_{X|D=B}(u)$$

The corresponding estimator is the following sample analogue:

$$\widehat{\Delta}_k^S = \frac{1}{N_B} \sum_{j:D_j=B} \widehat{\Delta}_k^S(X_j)$$

B Computation

The compute the detailed decomposition proposed in this paper, matrix formulation is recommended for reasons of efficiency. In what follows, $\otimes$ denotes the Kronecker product, $\circ$ the Hadamard product (element-by-element multiplication), $/\circ$ the Hadamard division (element-by-element division), and I_m is a column vector of ones of length m. Furthermore, $\mathbf{X}^A$ and $\mathbf{X}^B$ are the data matrices that contain the covariates of the subsamples of groups A and B, respectively. For the detailed structural effects, we first define:

$$\begin{aligned} A &= G(\mathbf{X}^B\beta^B) - G(\mathbf{X}^B\beta^A) \\ B &= \mathbf{X}^B \circ (I_{N_B} \otimes (\beta^B - \beta^A)') \\ C &= \mathbf{X}^B(\beta^B - \beta^A) \end{aligned}$$

Then, the $(1 \times K)-$vector of detailed structural effects is given by

$$\widehat{\Delta}^S = \frac{1}{N_B} I'_{N_B} \left[(I'_K \otimes A) \circ B \;/\circ\; (I'_K \otimes C) \right],$$

where the k-th element of $\widehat{\Delta}^S$ is the contribution of covariate k to the structural effect.

For the composition effect, computational efficiency can be increased considerably by expressing the inner summation with matrices. For each unit i in group A, define the following terms:

$$\begin{aligned} A_i &= G(\mathbf{X}^B\beta^A) - G(X_i\beta^A) \otimes I_{N_B} \\ B_i &= (\mathbf{X}^B - X_i \otimes I_{N_B}) \circ ((\beta^A)' \otimes I_{N_B}) \\ C_i &= [\mathbf{X}^B - X_i \otimes I_{N_B}]\beta^A \end{aligned}$$

In matrix notation, the inner summation can then be written as the $(1 \times K)-$vector

$$\widehat{\Delta}^X(X_i) = \frac{1}{N_B} I'_{N_B}[(I'_K \otimes A_i) \circ B_i / \circ (I'_K \otimes C_i)].$$

The unconditional contribution of covariate k to the composition effect is the k-th element of the $(1 \times K)$-vector

$$\widehat{\Delta}^X = \frac{1}{N_A} \sum_{i:D_i=A} \widehat{\Delta}^X(X_i).$$

Acknowledgements

I am grateful to three anonymous referees, Michael Gerfin, Blaise Melly, Stefan Boes, Kaspar Wüthrich, Steven Stillman, and various seminar participants for helpful comments on an earlier version of this paper.